

Unobtrusive Stress Assessment Using Smartphones

Alban Maxhuni*, Pablo Hernandez-Leal†, Eduardo F. Morales†, L. Enrique Sucar†, Venet Osmani‡, Oscar Mayora‡,

*Department of Health Technology, Technical University of Denmark, 2800 Kg. Lyngby, Denmark

almax@dtu.dk

†Instituto Nacional de Astrofisica, Optica y Electronica, Puebla, Mexico

{emorales, esucar, pablohl}@inaoep.mx

‡ Fondazione Bruno Kessler (FBK), Via Sommarive, 18, Povo 38123, Trento, Italy

{vosmani, omayora}@fbk.eu

Abstract—Stress assessment is a complex issue and numerous studies have examined factors that influence stress in working environments. Research studies have shown that monitoring individuals' behaviour parameters during daily life can also help assess stress levels. In this study, we examine assessment of work-related stress using features derived from sensors in smartphones. In particular, we use information from physical activity levels, location, social-interactions, social-activity and application usage during working days. Our study included 30 employees chosen from two different private companies, monitored over a period of 8 weeks in real work environments. The findings suggest that information from phone sensors shows important correlation with employees perceived stress level. Secondly, we used machine learning methods to classify perceived stress levels based on the analysis of information provided by smartphones. We used decision trees obtaining 67.57% accuracy and 71.73% after applying a semi-supervised method. Our results show that stress levels can be monitored in unobtrusive manner, through analysis of smartphone data.

Index Terms—Stress; smartphone-computing; behaviour monitoring; behaviour patterns; semi-supervised learning;

1 Introduction

WORK competition is becoming increasingly stronger, which together with the rapid economic transformation, have changed the dynamics of workplace environments. Due to these changes, enterprise employees are experiencing intense job insecurity, increased workloads, and longer working hours. All these factors engender work-related stress of different degrees, affecting the physiological and psychological functioning of employees. According to recent reports from the *European Agency for Safety and Health at Work* (EU-OSHA), stress was found to be the second most common work-related health problem across 27 Member states of the European Union (EU). Overall, 22% of EU employees reported work-related stress. This was accompanied by physical, psychological or social complaints, and was shown to be associated with the inability to bridge the gap with the requirements placed on them [1]. These patterns are similar in the U.S. where the National Institute for Occupational Safety and Health (NIOSH), reported that over 40% of employees suffer from workplace activities being "very or extremely" stressful [2].

Long-term exposure to stress can lead to serious health problems, ill physiological effects (e.g., fatigue, decreased sleep quality), behaviour changes (e.g., addiction, attention deficit), and social isolation issues (e.g., anger) [3, 4]. As a consequence, these negative effects have been shown to decrease well-being at the workplace and employees' work effectiveness. Moreover, long-term exposure to stress typically leads to job-burnout, a state that leads to mental and physical exhaustion [4]. Because of the previously mentioned reasons it is important to measure stress as a way of monitoring individual's well-being. However, stress is challenging to measure [2].

To date, physiological measurements and self-reported questionnaires are the most common methods used to infer work-related stress. In contrast, limited research has been

conducted within real-life workplace environments, as shown in the literature review, where a substantial number of studies recruit college students rather than employees of organisations. We aim to fill this gap by monitoring employees' realistic behaviour and its correlation with stress.

Smartphones have gained powerful sensing capabilities and the number of these devices is increasing, making them excellent candidates to be used for monitoring activities in working environments. Thus, the challenge is to use sensing capabilities of smartphones to detect stress-related behaviour of a person in an unobtrusive manner and take pre-emptive actions [5].

In this study we address the following research questions:

- Is there a correlation between the subjects' behavioural characteristics, extracted from smartphone sensor data, and their self-reported stress levels?
- Can limited labelled stress data be augmented with unlabelled sensor data to improve stress prediction accuracy?

Our approach focuses on behaviour changes that can be directly measured using smartphones: location changes, physical activities, social interactions and phone application usage. In processing data from 30 subjects we used: i) hierarchical clustering of perceived stress and identified the diversity and similarity of self-reported stress level within subjects and organisations; ii) decision tree classifiers to predict the stress level of the subjects using only data obtained from the smartphones (achieving a prediction accuracy of 67.5%); and iii) semi-supervised learning (SSL) methods [6] to address missing values (20.03%) in the self-reported questionnaires, increasing the accuracy to 71.73%; 67.8% for precision and 71.4% for recall in average, for all subjects.

In summary, the main contributions of the paper are:

- A comprehensive analysis of association between objectively measured data, including physical activity, location and social interaction with subjective self-assessment of work-related stress.
- Analysis of correlations between demographic information, specifically gender and temporal information with reported stress levels.
- Evaluation of our methods with a real-world dataset collected for a period of two months from 30 employees in two companies.

The rest of this paper is organized as follows: Section 2 summarises previous research with smartphones concerning the monitoring of stress in working environments. Section 3 provides information about participants demographics, how the data was collected and a detailed description of the attributes that were selected for predicting the perceived stress level of the subjects. Section 4 presents correlation analysis and results of our study. Section 5 describes the results regarding the prediction achieved by our models and discusses the potential benefits of the approach. Finally, conclusions and future research directions are given in Section 6.

2 Related Work

Several methods have tried to infer stress based on physiological signals, such as heart-rate variability, blood pressure, body temperatures and respiration [7, 8]. The work in [9] also provides short-term feedback to the users in everyday activities. However, the use of physiological sensors has reported several limitations: i) sensors require a large size to cover many signal types [7], ii) during the measurement period, movement is limited (i.e., because of skin conductance sensor) [10], and iii) sensors increase the discomfort [9] since they need to be carried at all times (in specific places on the body) in order to allow long-term continuous monitoring.

The miniaturization of wearable sensors has made possible to include a number of them in smartphones. In [11], the authors proposed a method for detecting stress based on speech analysis with accuracies of 81% and 76% for indoor and outdoor environments, respectively, using the vocal production of 14 subjects. However, in real-life activities (crowded environments) this approach may lead to misinterpretation of speech and therefore of emotion.

Research using smartphones for long-term stress monitoring [12, 13, 14] has proposed to collect many types of contextual data (e.g., physical activities, social activities and locations) that could help to infer stress from behaviour changes. In this line, Moodscope is a self-tracking system to help users manage their mood [15]. The system detects users' mood from smartphones usage data, such as e-mail messages, calls, SMS logs, application usage, web browsing histories and location changes. The authors reported an initial 66% accuracy of subjects' daily mood, improving to 93% after two months of training. It should be noted that the work on MoodScope focuses only on two narrow dimensions of mood, namely pleasure and activeness, which may explain the high prediction accuracy.

Table 1 summarises related works aiming at detecting the occurrence of stress using smartphones. It can be seen that there are few works (only two) that use only smartphones on realistic conditions with a reasonable number of subjects. The

TABLE 1: Main related works in stress detection showing the features used and details about the study.

Study	Items measured	Study details
Bauer et al.[16]	Smartphone (Location, bluetooth, phone-call and SMS)	Uncontrolled; 7 subjects, accuracy of 53%
Likwamw et al.[15]	Smartphone focused on mood (E-mails, Call and SMS logs, application usage, web browsing histories and location changes)	32 subjects; Reached overall accuracy from 66% to 93%
Sano et al.[17]	Wrist sensors and smartphone (Phone usage duration, phone calls and SMS logs)	Uncontrolled; 18 subjects; accuracy of 75%
Sano et al.[18]	Wearable sensors and smartphone (SMS logs, Location, Internet usage, E-mails, and overall phone usage)	Uncontrolled; 66 subjects; 1,980 days; Accuracy ranged from 67-92%.
Muaremi et al.[19]	HRV and smartphone (Phone- and SMS logs, Location, Audio Stress Response)	Uncontrolled, 35 subject from three different companies. Accuracy of 53% from smartphones and 61% after fusion
Bogomolov et al.[20]	Smartphone (Call logs, SMS, Bluetooth and weather)	Unconstrained, unknown ; 117 subjects. Overall accuracy of 72.39%

first one [15] is focused on mood, not in stress and the second one [20] is focused mainly on prediction and makes no analysis of perceived stress based on different demographic variables, social interaction, activity level, job-related variables, phone-usage, and location nor includes a correlation and regression analysis as we do in this paper.

In contrast to previous works, we aim at enabling measurement of relevant aspects from employees' behaviour in an unobtrusive way. Our trial was based on real fieldwork for a period of two months. In addition, we performed a systematic analysis for each sensor (physical activity level, social-interaction, locations and phone usage) and its relationship with stress levels. We provide a detailed overview of subjects behaviour patterns on a daily basis, taking into account their demographics (i.e., age, gender, education and family status). Using this information, we were able to investigate the association between objectively measured data with subjective self-assessment of work-related stress of various data groups, based on demographic information.

3 Dataset

Data was collected from a group of 30 subjects in the course of 8 weeks during the months of November and December. Participants were recruited on voluntary basis. A presentation was held on the premises of each organisation, with roughly the double of the final participants, outlining objectives and methods of the study. Most of the people that refused to participate on this study did it because they did not want to use another cellphone. Then, the interested participants were given the research pack containing the smartphone that was already configured with our data collection app. Except for the annual organisation-wide health screening, we have not carried out any additional clinical screening. Our data collection framework was based on a server-client architecture built around the Samsung Galaxy S3 mini 32GB smart-

phone¹. During the study, subjects used the smartphone in daily basis as their own phone (including working hours). There were no restrictions placed on users regarding the handling of their smartphones, so our analysis is framed under usual/realistic conditions. The application developed to collect data was running continuously in the background, starting automatically at 9am on working days (Monday-Friday) without any interaction from the user. In order to understand users' mood and stress levels, the app prompted users to fill in a questionnaire at three different times of the day: at 9am (at the beginning of the work hours), at 2pm (after lunch break) and at 5pm (at the end of the work hours). The questionnaires appeared automatically and the user had the option to answer the questions or snooze the questionnaire for later. The questionnaire consisted of 14 questions that were answered in around one minute. The questions relate to sleep quality, stress, and energy levels, among others (a more detailed description of the variables obtained from the questionnaires is provided later in Table 4). These questions were derived from two validated questionnaires, namely Oldenburg Burnout Inventory [21] and Profile of Mood States (POMS) [22]. Examples of screenshots of the questionnaire are shown in Figure 1.

Feature extraction was performed for two types of variables:

- Users' behaviour from sensors during work hours (*objective variables*).
- Scores obtained from the questionnaires (*subjective variables*).

Extracted data for each day was divided into two intervals: from 9am to 2pm, and from 2pm to 5pm, referring to the subjective variables (considered as ground truth) acquired from questionnaires.

3.1 Study Demographics

In total, 30 employees from two different organisations in Trento, Italy, were selected for the study. Table 2 provides a summary of employees' demographics characteristics. As it can be noted there is a fairly balanced mix of gender, age and education level, marital status and number of children among the subjects.

TABLE 2: Study demographics of the subjects in our study.

Variable	Characteristics	Nr. (%)
Gender	Male	18 (60.00%)
Education	High-school graduate	9 (30.00%)
	Bachelor degree	11 (36.67%)
	Graduate degree	10 (33.33%)
Age	26-30	5 (16.67%)
	31-40	18 (60.00%)
	>40	7 (23.33%)
Mean		37.46
Marital status	Married	15 (50.00%)
	Never married	15 (50.00%)
No. of children	None	17 (56.67%)
	1-2	10 (33.33%)
	3-4	3 (10.00%)

3.2 Subjective variables

We implemented POMS (Profile of Mood States) validated scale [22] as well as Oldenburg Burnout Inventory questionnaire [21] in the app. We derived 14 questions related to affect

1. We did not consider using other devices, such as smart watches, due to budget restrictions as well as limited availability at the time of the study.

of mood states, including, *Positive Affect* (PA) (e.g., cheerful, energetic, friendly) and *Negative Affect* (NA) (e.g., tense, anxious, sad, angry), and the rest measures disengagement from work. The PA, NA and disengagement from work items were presented in mixed order.

In our study, each question had five response alternatives, which assessed five stress-related factors on a scale ranging from 1 (absolutely agree) to 5 (absolutely disagree). The answers were stored on the mobile device and were segmented into three regions: "low" or "poor", score < 3; "moderate" or "fair", score = 3; and "high" or "sufficient", score > 3.

The first section of the questionnaire collected information about occupational health outcomes of the participants: i) job induced stress, ii) job-control, iii) job-demand and iv) energy perceived during working days. The second section contained several questions to measure mood: the existence of tensions and pressures growing out of job requirements, feelings of anxiety, cheerfulness, friendliness, sadness, angeriness, and quality of sleep.

3.3 Objective variables

The second type of data was collected from sensors embedded on the smartphones. From the analysis presented in Section 2 we concluded that 4 categories were needed to perform a proper assessment of subjects stress: physical activity, location, social interaction and social activity. From these categories we extracted 19 features using 8 sensors, as shown in Table 3. In the following sections we discuss each category in detail.

TABLE 3: Objective variables divided in four categories. Sensors and features extracted from smartphone usage on every subject in the study.

Category	Sensors	Features
1. Physical Activity Level	Accelerometer	1) 3-axis Magnitude 2) Variance Sum [23]
2. Location	Cellular	3) CellID and LACID (Number of clusters - DBSCAN) [24]
	WiFi	4) Access Points (Number of clusters - DBSCAN) [24])
	Google-Maps	5) Latitude and Longitude (Number of clusters - DBSCAN) [24], Haversine [25])
3. Social Interaction	Microphone	6) Proximity based on verbal interaction (Pitch [26], Mel-MBSES [27] and WiFi [28])
	Phone Calls	7) Number of Incoming Calls 8) Number of Outgoing Calls 9) Number of missed Calls 10) Duration of Incoming Calls 11) Duration of Outgoing Calls 12) Most common Contact-Calls 13) Number of Incoming SMSs 14) Number of Outgoing SMSs 15) Length of Incoming SMSs 16) Length of Outgoing SMSs 17) Most common Contact-SMS
	SMS	18) Number of used applications (Social and system) 19) Duration of used applications (Social and system)
4. Social Activity	App usage	

3.3.1 Physical Activity Level - (pACL)

On the one hand, psychological stress has been reported as a factor in reducing frequency, intensity, and duration of physical activity [29]. On the other hand, research studies have acknowledged physical activity as a psychological de-stressor. We wanted to investigate the association between



Fig. 1: Examples of screenshots of the graphical user interface on the smartphone of the survey questionnaire. Questions shown were as follows, in English (from the left, screenshot 1) How well did you sleep last night?, What is your level of stress? (screenshot 2). This activity is a challenge:, This activity is something I'm good at:, I would prefer to do something else: (screenshot 3) Anxiety, Efficiency, (screenshot 4) Would you like to continue with the questionnaire?

objectively measured physical activity and perceived psychological stress. For this study, we captured 3-axial linear acceleration continuously at a rate of 5Hz. Similar to the work in [23], we measured the variance sum (varSum) of 26 seconds (non-overlapping fixed length windows of $n=128$ samples) accelerometer readings, providing the activity levels of *high*, *low*, and *none* using the magnitude of the signal.

We define three ranges of *percentage of physical activity level (pACL)* as follows: *high*-(h) when $\text{varSum} \geq 7$, *low*-(l) when $3 \leq \text{varSum} \leq 7$, and *none*-(n) when $\text{varSum} < 3$; using Equation 1:

$$\text{pACL}_{(h,l,n)} = \frac{\text{Number of High Activities (h)}}{\text{Total Classified Activities (h,l,n)}} \times 100\% \quad (1)$$

3.3.2 Location

We analyse locations of subjects with the focus on understanding frequent locations changes during working hours. We retrieved: (i) the list of WiFi networks available with their respective Basic Service Set Identifier (BSSID) address, (ii) cell tower locations (Cell Identifier (CID), Locations Area Cell Identifier (LAC-ID)) and (iii) Google Maps locations information (latitude, longitude).² Using the location information we clustered locations from each source using the DBSCAN algorithm [24], which is an algorithm mainly used for clustering spatio-temporal locations. For Google location information, we clustered locations with maximal diameter of 300 meters (using latitudinal and longitudinal coordinates and the Haversine distance equation [25]) where the subjects stay for more than 15 minutes and measured the amount of locations in each day. For cell tower information and WiFi networks we clustered location information on an hourly basis. We compared locations every hour counting +1 when different clusters appear with respect to the previous hour.

3.3.3 Social Interaction (SI)

We investigate the effects of stress on social behaviour derived from continuously recorded and classified human voice (from smartphone's microphone) in real working environments. Another important social aspect are the employees phone conversations and SMS logs.

² We have intentionally not used the GPS sensor, in order to preserve the battery life of the smartphones. However, we use Google Locations Services which use information from wi-fi networks, cell towers, accelerometers, IP addresses, etc., to find the location of the user.

We measured two aspects of social interaction:

- **Speaker Recognition:** We use the microphone embedded on the smartphones for better and accurate recognition of verbal interactions (namely social-interactions).³ We extracted two main audio features: Pitch [26] and Mel-MultiBand Spectral Entropy Signature (Mel-MBSES) [27] to obtain a higher accuracy in speech activity recognition. We built a Support Vector Machine (SVM) using Mel MultiBand Spectral Entropy Signature (Mel-MBSES) coefficients trained on frames coming from 3 minutes of voiced data (positive vector) and 3 minutes of background data (negative vector). We sampled audio frequency at 8000Hz and set a frame every 256 samples where we calculated Pitch and Mel-MBSES features for each frame, then each frame is labelled either as human voice or not human voice. Approximately every 0.7 seconds (7 out of 30 frames) a frame must be detected as voice in order to indicate voice activity in that audio segment. We measured percentage of social-interaction based on the total duration (hourly, daily, weekdays) of conversations as shown in Equation 2:

$$\text{Social-Interaction} = \sum_{i=1}^n \frac{\text{true - classified}}{\text{total - classified}} \times 100\% \quad (2)$$

- **Phone-Call and SMS behaviour:** We considered phone calls in terms of: *number*, *duration* and *most frequent number* (on a daily basis) of incoming, outgoing and missed. For SMSs, we measured the *number* and *length* (incoming and outgoing) of messages. In order to find the most common called/calling ID in each interval (9am-2pm and 2pm-5pm) we used $\text{argmax}_{(\text{Call})} = \sum_{i=1}^n \text{countmax}(\text{CallID})$ and $\text{argmax}_{(\text{SMS})} = \sum_{i=1}^n \text{countmax}(\text{SMSID})$ for most frequent Calls and SMSs, respectively. In order to remove ties among IDs that have the same number of calls, we proposed a scoring model, *Score*, for both calls and SMSs:

$$\text{Score}_{(\text{Call})} = \frac{\text{duration}(\text{CallID})}{\text{countmax}(\text{CallID})};$$

$$\text{Score}_{(\text{SMS})} = \frac{\text{length}(\text{SMSID})}{\text{countmax}(\text{SMSID})}.$$

³ The application did not store conversations. In situations like incoming and outgoing phone calls, accepting or dialling a phone call; we stop the recognition service on the phone and restarted service after the phone-calls ended.

3.3.4 Social-Activity

Finally, another aspect that may have impact on the stress levels is application usage. To capture this information, each time an employee used an app, our software stored the event together with the duration and time-stamp. We extract the number of application used per interval and their duration. Applications were divided in two categories:

- System apps: pre-installed apps like Camera or Calendar, Web-browsing, E-Mail client.
- Social apps: such as Viber, WhatsApp, Facebook, Skype and other user downloaded apps (e.g., games and other entertainment apps).

4 Analysis of information

Using the features presented in Section 3: (i) we applied hierarchical clustering (Section 4.1), and (ii) correlation analysis (Section 4.2). We have used a number of scales to asses various aspects of participants' states, including a Likert rating scale to measure stress levels. Furthermore, 14 questions were derived from two validated scales, namely Oldenburg Burnout Inventory [30] and Profile of Mood States (POMS) [22].

Table 4 provides the overall response rates of the completed questionnaires. We obtained 1455 completed questionnaires, which represented a response rate of 79.97%. In this study we included only self-reported questionnaire items obtained at \sim 2pm and \sim 5pm, since we are interested in exploring the relation of stress, mood, and job-performance with respect to the objective variables measured in the previous working hours. It can be seen that employees perceived a *moderate* (35.15%) to *high* (22.18%) stress level throughout the entire monitoring period. Almost all of them (29 out of 30) reported that at some point their job tasks and job responsibilities were highly demanding (50.58%) throughout the entire monitoring period (marked with red-color in Table 4). This is important since prolonged exposure to certain job-demands has been shown to lead employees to variety of health issues, such as mental and physical disorder [4]. In response to work-related stress, 19 employees felt themselves *High - Tensed* at some point of the study, 18 respondents felt *High - Anxious*, 11 of respondents have reported *High - Angriness* (5.67%), which shows that a large group of subjects showed negative moods. Finally, a relevant physical reaction to stress is a *Poor - Sleep Quality*, which was reported by 24 of the respondents.

4.1 Diversity and Similarity of Stress Level within Subjects

Hierarchical clustering was used to analyse the participants self-reported stress on a daily basis. We used Ward's method [31] using the half-square euclidean distance⁴ between subjects.

Table 5 provides an overview of the clustering results based on gender. Cluster analysis yielded 3 distinct clusters (C1, C2 and C3) which represent *low*, *moderate*, and *high* stress levels. Note that women show a uniform distribution across stress levels and men showed slightly more subjects with low stress.

4. Euclidean distance is always greater than or equal to zero. Measurements would be ≈ 0 for identical subjects and ≈ 1 for subjects that show less similarity.

TABLE 4: Subjective variables: overall percentage of self-reported questionnaires (exhaustion and disengagement from work) by perceived level (High, Moderate, Low) and number of subjects.

Variable	Level	Nr. Responses(%)	Nr. Subjects
Perceived Stress	<i>High</i>	325 (22.18%)	27
	<i>Moderate</i>	515 (35.15%)	30
	<i>Low</i>	625 (42.66%)	30
Perceived Job-demand	<i>High</i>	741 (50.58%)	29
	<i>Moderate</i>	357 (24.37%)	30
	<i>Low</i>	367 (25.05%)	24
Tensed	<i>High</i>	118 (8.06%)	19
	<i>Moderate</i>	280 (19.11%)	28
	<i>Low</i>	1067 (72.83%)	30
Cheerful	<i>High</i>	274 (18.70%)	28
	<i>Moderate</i>	756 (51.60%)	30
	<i>Low</i>	435 (29.70%)	30
Angry	<i>High</i>	83 (5.67%)	11
	<i>Moderate</i>	186 (12.70%)	5
	<i>Low</i>	1196 (81.63%)	30
Sleep quality	<i>Sufficient</i>	886 (60.48%)	30
	<i>Fair</i>	313 (21.37%)	28
	<i>Poor</i>	266 (18.15%)	24
Perceived Job-control	<i>High</i>	612 (41.77%)	30
	<i>Moderate</i>	604 (41.23%)	30
	<i>Low</i>	249 (17.00%)	27
Perceived Energy	<i>High</i>	357 (24.37%)	28
	<i>Moderate</i>	756 (51.60%)	30
	<i>Low</i>	352 (24.03%)	28
Anxious	<i>High</i>	128 (8.74%)	18
	<i>Moderate</i>	279 (19.04%)	3
	<i>Low</i>	1058 (72.22%)	30
Friendly	<i>High</i>	463 (31.60%)	27
	<i>Moderate</i>	692 (47.23%)	30
	<i>Low</i>	310 (21.16%)	29
Sad	<i>High</i>	28 (1.91%)	10
	<i>Moderate</i>	112 (7.65%)	30
	<i>Low</i>	1325 (90.44%)	12

We also performed clustering within the organisations, which is shown in the Table 6. For example, in organisation A (an IT organization), all women (4) showed high stress levels. In contrast, in organisation B (a social support organization), half of the women showed low stress and half of the women showed moderate stress levels. Again, in this company, there are slightly more men with low level of stress.

Finally, we clustered self-reported stress changes within intervals (9am-2pm and 2pm-5pm) as shown in Table 7. For example, low \longleftrightarrow moderate, means that subjects in the clusters showed low stress levels in the first interval and then changed to moderate in the second interval or that moderate changed to low. In this case, 23.33% of the subjects showed at least a high level of stress in their daily activities ($high \longleftrightarrow moderate$ or $high \longleftrightarrow high$) and 2/3 of the subjects (63.33%) showed levels between moderate and high. It is important to note that employees did not perceive drastic changes of stress, such as $low \longleftrightarrow high$.

As a summary of this first set of experiments: (i) there is a slight bias in men towards lower levels of stress in their working environments, (ii) there is a clear difference between stress levels in companies, where an IT company showed higher stress levels than a social support company, (iii) about 2/3 of the employees perceived moderate to high stress and 23.33% perceived high stress, and (iv) there were no drastic changes between levels of stress.

Now we present a more detailed analysis for each category of objective variables and its relation with mood, and specifically with perceived stress levels.

TABLE 5: Perceived stress level from dendrogram analysis by gender. Three major clusters can be noted based on perceived level of stress.

Cluster (Stress-Level)	Men (Nr./%)	Women (Nr./%)
C1 (low < 3)	7/18 (38.89%)	4/12 (33.33%)
C2 (moderate = 3)	6/18 (33.33%)	4/12 (33.33%)
C3 (high > 3)	5/18 (27.78%)	4/12 (33.33%)

TABLE 6: Perceived stress level from dendrogram analysis by gender within organisations. Three major clusters can be noted based on perceived stress levels.

Cluster (Stress-Level)	Org. A	Org. B
C1 (low<3)	Men: 3/12 (25.00%)	4/6 (66.67%)
	Women: 0/4 (0.00%)	4/8 (50.00%)
C2 (moderate=3)	Men: 4/12 (33.33%)	2/6 (33.33%)
	Women: 0/4 (0.00%)	4/8 (50.00%)
C3 (high>3)	Men: 5/12 (41.67%)	0/6 (0.00%)
	Women: 4/4 (100.00%)	0/8 (0.00%)

TABLE 7: Perceived stress level from dendrogram analysis by response intervals ([9am-2pm], [2pm-5pm]). Three major clusters can be noted based on perceived level of stress and transition of perceives stress into intervals ([9am-2pm] and [2pm-5pm]).

Cluster (Stress-Level)	Intervals
C1 low↔low; low↔moderate	11/30 (36.67%)
C2 moderate↔low; moderate↔moderate	12/30 (40.00%)
C3 high↔moderate; high↔high	7/30 (23.33%)

4.1.1 Physical Activity Levels

Table 8 presents the overall percentage of physical activity level with respect to perceived stress level (*High*, *Moderate*, and *Low*) compared with demographic characteristics (age, gender, education, marital status, number of children and organisations). Activity levels were normalized for each interval (9am-2pm and 2pm-5pm) or for a complete day. Based on this analysis we can see that:

- pACL during lower perceived stress times was associated with higher activity (19.65% of activity). In contrast, a high perceived stress showed less activity (16.43%).
- Subjects were more active during the second interval (2pm-5pm), with 20.14% pACL compared to 17.06% in the first interval.
- Male subjects had higher physical activity levels than female subjects.
- Following age, education level, and marital status, participants that reported *high* and *moderate* stress levels were associated mostly with lower pACL than when they have *low* stress.
- The age group of (≥ 40) showed more activity level than the rest when they perceived high stress level.
- The group of married subjects showed more activity than the never married group no matter their perceived stress level.
- Subjects from company B report less stress and were physically more active than subjects from company A.

Furthermore, separating overall activities into working days (see Table 9) show that men have a higher pACL on Mondays when they perceive high stress, while the rest of the week they showed higher percentage of activity with low perceived stress. In contrast to women that showed more activity with low levels of perceived stress at the beginning

of the week and more activity with moderate levels of stress at the end of the week.

Table 10 shows mean scores on perceived job-demand, perceived job-control, perceived stress, and perceived energy for the respondents. From the table we can observe that low perceived job-demand was associated with higher physical activity level (22.3%) for male participants. In contrast, women showed increased activity levels when they perceive high job-demands. Similarly, men participants with higher pACL perceived higher energy. In contrast, women with higher pACL showed lower energy. Also men with lower pACL showed higher percentage of sleep-quality. In summary, men and women showed opposite results in terms of perceived job-demand and energy with respect to their activity levels throughout the day.

4.1.2 Social Interaction

In contrast to our previous work [32], where we explored the correlation of total amount of verbal interaction per day with self-reported stress, in this study we expand that analysis, since now we explore the distribution of the verbal-interaction in an hourly basis and working intervals.

Table 11 presents a summary of social-interaction levels. Some findings are the following:

- Both male and female subjects showed higher social interaction in moments of high stress.
- However, analysing this data by age group we observe that older (and married) employees showed the opposite behaviour, they increased their social interaction during low levels of perceived stress.
- There is in general more social interaction in the afternoons than in the mornings.
- Subjects in organisation A showed higher social interaction than those in organisation B.

We explored further these measurements. We depict in Table 12 and Table 13 social-interaction as percentage in an hourly basis in a day, per day of the week, per hour within organisations and per day of the week by gender.

- A notable result is an homogeneous behaviour of social interaction across stress levels (Table 12), with higher interaction in the morning for moderate perceived stress and a higher interaction in the afternoon for high perceived stress.
- Another homogeneous behaviour is shown across organisations, where people decrease their social interaction near lunch time (12-13 hrs), see Table 13.
- The social interaction varies with the perceived stress during the week. When subjects perceive high stress, social interaction increases on Tuesdays and Wednesdays and then decreases on Thursdays and Fridays, see Table 12.
- With respect to gender, both men and women showed a more stable social interaction across the weekdays .

Using the self-reported stress level, we were able to compare the phone activeness from 5767 phone calls and 5911 SMSs.⁵ (see Tables 14 and 15). From these tables it can be seen that the number of phone-placed *Outgoing*, phone received *Incoming* and missing calls, was higher when subjects perceive less stress.

5. All marketing SMSs or responses from the GSM operators were excluded in this study.

TABLE 8: Overall average percentage of physical activity level (pACL) by intervals (9am-2pm and 2pm-5pm) and perceived stress level (SL) [High, Moderate, Low].

Distribution of pACL by (Gender, Age, Education, Marital Status and organisation)	pACL [9-2]	pACL [2-5]	High (SL)	Moderate (SL)	Low (SL)
– Male	18.03	21.34	16.29 (*)	16.68	23.60
– Female	15.66	18.74	10.57 (**)	15.37	18.89
– 26-30 (28.6±1.95)	12.89	15.48	12.45	13.65	17.83
– 31-40 (35.33±2.4)	17.50	21.00	12.87	16.22	21.97
– >40 (49±2.52)	18.69	21.66	17.61	18.20	21.90
– High school graduate	17.01	21.40	16.84	16.84	18.77
– Bachelor degree	19.22	23.52	11.70	17.48	29.19
– Graduate degree	14.78	15.54	12.64	14.86	16.51
– Married	20.51	25.48	17.71	19.53	26.73
– Never married	13.36	14.78	10.23	13.39	16.31
– Organisation A.	12.17	15.50	12.21	10.77	17.33
– Organisation B.	22.45	25.49	18.39	23.93	24.21
– Overall (Mean±SD) of pACL (%)	17.06 (±12.01)	20.14 (±13.12)	16.43 (±16.42)	16.46 (±12.30)	19.65 (±12.85)

(*) 16/18 - male subjects perceived high stress.

(**) 11/12 - female subjects perceived high stress.

TABLE 9: Overall average percentage of activity level (Mean±SD) during working days and perceived level (SL) of stress (H-High, M-Moderate, L-Low) by gender.

	Men			Women		
	H (SL)	M (SL)	L (SL)	H (SL)	M (SL)	L (SL)
Monday:	24.3±22.2	16.2±16.2	21.6±18.0	12.3±12.1	13.0±7.0	21.6±22.4
Tuesday:	10.0±6.5	17.5±16.6	22.2±14.4	6.2±3.1	12.3±6.3	16.5±7.7
Wednesday:	18.0±19.8	19.8±18.3	22.5±18.5	12.6±8.4	13.2±7.7	14.6±7.6
Thursday:	19.0±20.7	20.7±18.6	24.3±18.7	9.6±8.0	17.9±12.4	14.3±13.9
Friday:	14.9±17.3	15.9±19.6	20.4±19.5	11.4±12.4	17.7±13.6	13.8±8.0

TABLE 10: Overall average percentage of activity level (Mean±SD) by job-demands, job-control, energy and sleep-quality perceived level (PL) with respect to gender.

	Men			Women		
	H (PL)	M (PL)	L (PL)	H (PL)	M (PL)	L (PL)
Job-Demand (%)	20.0 ± 16.7	17.9±14.8	22.3 ± 22.1	16.1 ± 6.0	13.1±8.4	11.7±6.2
Job-Control (%)	19.0 ± 14.6	18.0±14.8	18.9±16.0	14.2±6.5	16.7 ± 6.2	13.3 ± 8.7
Energy (%)	23.7 ± 15.6	19.9±14.8	17.8±14.6	14.1±9.7	15.3±6.1	16.8 ± 9.7
Sleep-Quality (%)	20.9±14.3	22.1±15.8	22.7 ± 17.4	16.6 ± 6.9	15.9±6.8	15.8±6.9

TABLE 11: Distribution of social-interaction (SI) by response intervals ([9am.-2pm.],[2pm. - 5pm.]) and stress-level (SL)

Distribution of SI by Gender, Age, Education, Marital Status, Organisation	SI [9-2]	SI [2-5]	High-(SL)	Moderate-(SL)	Low-(SL)	Nr. Employees
– Male	25.67	28.75	27.88 (*)	27.54	25.74	18
– Women	20.17	23.83	22.88 (**)	22.72	19.79	12
– 26-30 (28.6±1.95)	25.46	29.53	28.57	26.02	26.44	5
– 31-40 (35.33±2.4)	22.96	26.61	24.90	26.67	22.34	18
– >40 (49±2.52)	22.73	24.84	22.97	22.54	24.32	7
– High school graduate	20.63	25.09	22.97	26.95	26.22	11
– Bachelor degree	24.23	28.16	29.49	26.28	23.12	10
– Graduate degree	25.30	26.94	22.81	23.37	21.24	9
– Married	21.75	25.02	22.91	21.92	23.56	15
– Never married	24.68	28.07	27.61	28.45	22.89	15
– Organisation A.	26.40	30.41	27.96	29.64	25.67	16
– Organisation B.	20.07	22.49	18.20	20.21	21.80	14
– Overall (Mean±SD) of SI (%)	23.61 ± 10.53	26.93 ± 11.04	23.47 ± 11.02	24.58 ± 10.47	25.28 ± 11.67	30

(*) 16/18 - male subjects perceived high stress.

(**) 11/12 - female subjects perceived high stress.

TABLE 12: Distribution of social-interaction (SI) by day of the week, stress-level (SL) and gender.

Stress Level	Mon	Tue	Wed	Thur	Fri
High	20.7±14.0	26.4±15.9	27.5±17.0	21.7±15.6	25.8±14.5
Moderate	25.4±13.9	23.2±18.0	23.4±14.4	31.5±20.1	26.5±11.6
Low	26.1±13.4	23.9±14.4	23.8±16.5	27.1±13.9	23.7±15.1
Gender	Mon	Tue	Wed	Thur	Fri
Male	28.1±12.4	26.2±12.1	25.3±11.6	27.0±10.8	26.0±9.2
Women	23.1±12.9	18.9±13.2	21.4±11.13	21.4±11.1	21.1±11.0

We also analysed the duration and length of calls and SMSs:

- In stress-full days, outgoing calls have in average shorter duration.
- Longer duration of incoming calls were associated with high perceived stress level.
- Almost in all cases a high number (and length) of incoming and outgoing SMSs were also related to *high* stress.
- Analysing the conversations by weekdays, high perceived stress was associated with longer duration of incoming calls and the length of incoming SMSs, which in contrary to duration of outgoing calls and length of outgoing SMSs, is lower when the employees perceive high stress. Similarly, having high job-demands was associated with lower duration of

TABLE 13: Distribution of social-interaction (SI) by working hours, stress-level (SL) and organization.

Stress Level	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm
High	20.8±11.7	24.0±13.3	23.5±15.1	22.9±14.5	22.2±15.5	26.0±15.8	25.3±17.0	25.9±16.4	23.8±18.1
Moderate	29.0±17.2	25.0±14.3	27.8±13.6	21.8±13.1	19.1±10.7	24.0±12.9	25.5±14.7	27.4±15.1	18.5±18.2
Low	25.1±14.9	25.3±14.3	27.6±17.1	22.7±13.6	21.6±10.3	26.7±14.3	30.8±18.5	26.0±15.2	23.1±22.8
Organization	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm
A	29.1±11.0	27.1±13.7	28.5±13.5	22.8±11.5	21.6±10.4	28.9±10.5	31.1±11.5	31.6±12.9	5.8±24.8
B	20.9±11.8	22.9±11.1	22.9±11.9	21.2±8.7	18.6±6.8	22.5±9.6	24.2±12.9	22.8±8.1	18.8±13.3

TABLE 14: Number of phone-calls by perceived stress level (SL).

	Nr. Phone Calls	High SL	Moderate SL	Low SL
Incoming:	1696 (100%)	355 (20.9%)	511 (30.1%)	830 (48.9%)
Outgoing:	2912 (100%)	547 (18.7%)	839 (28.8%)	1526 (52.4%)
Missing:	1159 (100%)	220 (18.9%)	405 (34.9%)	534 (46.1%)

TABLE 15: Number of SMSs by perceived stress level (SL).

	Nr. SMS	High SL	Moderate SL	Low SL
Incoming:	3767 (100%)	1067 (28.3%)	801 (21.2%)	1899 (50.4%)
Outgoing:	2144 (100%)	697 (32.5%)	710 (33.1%)	737 (34.3%)

phone-calls and length of SMSs in all categories.

Moreover, figure 2(a) depicts the frequency of the most common contact for phone calls and SMSs (blue line figure 2(b)) for every subject. From these figures we note a higher frequency of phone-calls and SMSs with the most common contacted number when they perceive high stress levels (average frequency of most frequent contacts shown in red line). In contrast, in low and moderate stress the frequency of phone-calls is lower in average. These results show that a higher frequency of the phone-calls and SMSs is correlated with stress levels during working times.

4.1.3 Location Changes

To analyse location changes we measured the number of clusters obtained from different locations throughout the entire monitoring interval (see Table 16). From all three sources it is evident that, overall, subjects tend to reduce visiting different places or going further away from work environments when they perceive high stress levels.

4.1.4 Application Usage

Another source that provides information relevant to subjects daily activities at work is the usage of the smartphone applications. Table 17 shows that in periods with very low stress subjects tend to use longer times the smartphone (both with social and system applications). This also seems a good indicator for identifying perceived stress levels.

In summary, from these results we can draw the following conclusions:

- Activity levels changed with perceived stress and with weekdays.
- There is an opposite behaviour of activity levels in male and female in terms of job-demand and energy.
- There is more social interaction with higher stress levels except for older people that show an opposite behaviour.
- There is more social interaction during the afternoons.

TABLE 16: Overall number of clusters obtained from location using the DBSCAN algorithm by perceived stress level (SL). Descriptive statistics (Mean±SD) provide information of overall number of clusters retrieved from the 30-subjects throughout the entire monitoring period.

Locations	Clusters 9am-5pm - Nr. (Mean±SD)	High- (SL) - Nr. (Mean±SD)	Moderate- (SL) - Nr. (Mean±SD)	Low- (SL) - Nr. (Mean±SD)
Cell:	1383 (1.05±0.38)	230 (1.01±0.39)	349 (1.07±0.40)	527 (1.05±0.33)
WiFi AP's:	2663 (1.40±1.38)	486 (1.42±1.41)	742 (1.49±1.35)	961 (1.55±1.39)
Google Maps:	628 (0.48±0.78)	143 (0.63±1.01)	158 (0.48±0.90)	234 (0.46±0.85)

TABLE 17: Overall number/duration (seconds) of phone application usage by perceived stress level (SL). Descriptive statistics (Mean±SD) provides overall usage of applications from 30 subjects during the entire monitoring period.

Perceived Stress Level	Frequency System- Apps - Nr. (Mean±SD)	Frequency Social- Apps - Nr. (Mean±SD)	Duration System- Apps - Nr. (Mean±SD)	Duration Social- Apps - Nr. (Mean±SD)
High	5531 (24.0±26.1)	357 (3.5±4.0)	48445 (211.0±157.2)	4621 (45.3±52.5)
Moderate	7823 (25.2±28.5)	508 (4.0±4.2)	57607 (185.2±153.3)	7420 (57.0±73.2)
Low	13787 (31.0±28.2)	966 (4.3±4.2)	88782 (197.2±150.3)	9582 (42.3±65.2)

- There is an increased level in social interaction by women towards the end of the week.
- There is a very different social interaction among employees of different companies. Curiously the company with higher stress levels also have higher percentages of social interaction.
- There are shorter outgoing calls and longer incoming calls during high stress levels.
- People use much more their smartphones during lower perceived stress levels.

The next section explores the correlation between objective and subjective variables, as well as the importance of each variable for mood prediction.

4.2 Correlation between objective and subjective data

We conducted two types of correlation analysis to investigate the association between four factors: perceived stress, negative-mood, positive-mood, and overall mood score. Emotions were divided in two categories: negative-mood (tense, stress, angry, anxious and sad) and positive-mood (friendly, energetic, cheerful and being good at current activity). An overall score derived from both types of emotions was ob-

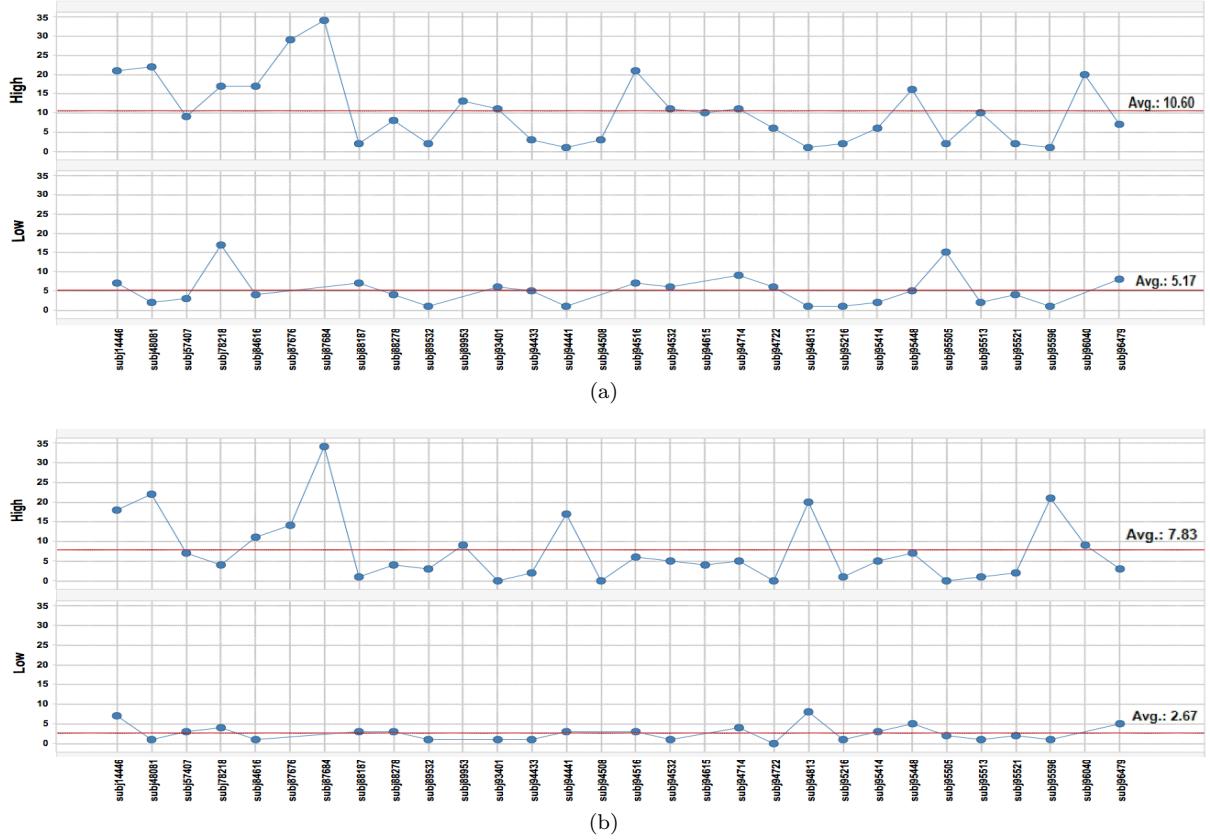


Fig. 2: Frequency of the most common contact (a) calls and (b) SMSs for each subject by perceived stress level (High/Low).

tained by subtracting negative mood scores from positive scores.

First, we performed two-tailed Pearson correlation tests between objective and subjective variables. The Pearson correlation coefficient ρ was used, with a statistically significant value when $\rho < 0.05$ (*) and $\rho < 0.01$ (**). Table 18 summarizes the results.

- For stress level, physical activity ($r = -0.153$, **, $N=1465$), number of WiFi ($r = -0.087$, **, $N=1456$) and cellular location ($r = -0.070$, *, $N=1456$), number and duration of outgoing phone calls ($r = -0.098$, *, $N=1120$), number and length of SMS responses ($r = 0.090$, *, $N=505$), and number of system apps ($r = -0.129$, **, $N=1292$) obtained statistically significant correlations.
- In particular, for missing calls we expected to have correlation with different factors. However, it was shown to have a weak correlations with the stress factor.
- Negative emotions show high correlation with physical activity, number of system apps, social interaction information and social activeness (number of incoming and outgoing phones calls, and outgoing SMSs).
- Social interaction information, the number of outgoing calls and the use of social applications showed high correlation with positive mood scores.
- Information from the duration of social applications and location information had no statistically significant correlation with negative emotions. This is interesting because these same two variables had high correlation with positive emotions. Similarly the number of incoming calls had no statistically significant correlation with positive emotions.

but is highly correlated with negative emotions.

TABLE 18: Pearson correlation between objective variables and perceived stress level, negative mood score, positive mood score, and total mood score (TMS).

Objective Variables	Stress Level	Negative Mood	Positive Mood	TMS
pACL	-0.153**	-0.112**	0.071**	0.116**
Cellular Location	-0.070 *	-0.070*	0.033	0.065*
Google-Maps Loc.	0.051	0.017	0.079*	0.033
WiFi Location	0.087**	0.039	-0.120**	-0.093**
Social Interaction	0.032	0.059*	-0.142**	-0.119**
Nr-Out-Calls	-0.980**	-0.112**	0.083**	0.121**
Nr-In-Calls	-0.005	-0.090**	-0.019	0.05
Missed-In-Calls	-0.006	-0.023	-0.012	0.009
Dur.-Out-Calls	-0.098**	-0.097**	0.101**	0.123**
Dur.-In-Calls	0.037	-0.034	0.091*	0.074*
Nr-SMS-Out	0.090**	-0.071*	0.004	0.05
Nr-SMS-In	0.006	-0.012	-0.044	-0.016
Length-SMS-Out	-0.154**	-0.153**	0.106*	0.156**
Length-SMS-In	0.013	-0.028	0.088*	0.069
Dur.-App-System	0.008	-0.021	-0.024	0.001
Dur.-App-Social	0.067	0.067	-0.218**	-0.161**
Nr-App-System	-0.129**	-0.181**	0.194**	0.228**
Nr-App-Social	-0.060	-0.040	-0.004	0.024

Significant at levels: $\rho < 0.05$ (*); $\rho < 0.01$ (**).

5 Prediction of perceived stress

Predicting perceived stress of the user can be seen as a classification problem. In this case, the attributes correspond to each feature related to the objective variables and the class to predict is the self-reported stress level (low, moderate, high). Since we are interested in analysing behaviour changes or patterns that may appear in daily activities, we used decision trees [33] which can be easily understood. Our approach was

to build a decision tree for each subject of the study, with the idea of analysing individual behaviours and models.

We performed classification of the stress variable using the C4.5 algorithm [33] and 10-fold cross validation for each user. Table 19 presents the classification accuracy and the average precision, recall and f-measure for stress level for the 30 subjects. In average the accuracy obtained was 67.57%. However, as previously mentioned, the dataset contained 20% of missing data. This is an important portion which can be exploited with a semi-supervised learning (SSL) technique.

TABLE 19: Stress prediction using decision trees before and after applying a semi-supervised learning approach.

Subjects	Supervised	Semi-Supervised	Overall Increase
Acc. Mean \pm SD:	67.57% (\pm 15.60%)	71.73% (\pm 15.25%)	4.20% (\pm 9.52%)
Labelled Data:	79.97% (1465/1832)	94.00% (1722/1832)	14.03%
Precision (%):	65.4%	68.9%	
Recall (%):	68.9%	73.0%	
F-Score (%):	66.0%	70.0%	

TABLE 20: Objective variables - percentage Variable of Importance (VI) and Weight by Tree Importance (WTI) using decision tree classifier.

Objective Variables	VI (%)	WTI (Mean \pm SD)
pACL	13.3%	86.7 \pm 18.8
Cellular Location	40.0%	18.0 \pm 12.6
Google-Maps Location	13.3%	14.5 \pm 12.4
WiFi Location	60.0%	39.2 \pm 26.5
Social Interaction	26.7%	45.4 \pm 22.8
Number-Outgoing-Calls	100.0%	15.3 \pm 11.0
Number-Incoming-Calls	36.7%	14.6 \pm 13.8
Missed-Incoming-Calls	23.3%	7.5 \pm 7.8
Duration-Outgoing-Calls	20.0%	13.1 \pm 9.7
Duration-Incoming-Calls	23.3%	10.8 \pm 9.9
Number-SMS-Outgoing	40.0%	8.2 \pm 5.6
Number-SMS-Incoming	43.3%	8.4 \pm 5.0
Length-SMS-Outgoing	16.7%	6.7 \pm 5.7
Length-SMS-Incoming	33.3%	8.7 \pm 9.4
Duration-Application-System	17.0%	67.0 \pm 28.2
Duration-Application-Social	16.7%	32.9 \pm 22.6
Number-Application-System	33.3%	31.8 \pm 20.7
Number-Application-Social	16.7%	14.7 \pm 10.0

5.1 Semi-supervised learning

For this study we consider one of the most common methods of SSL that uses a single classifier called Self-Training [34]. This method selects the most confident unlabelled points, together with their predicted labels and adds them to the training set. In each iteration the newly high-confidence (>80%) labelled instances are added to the original labelled data. The classifier is re-trained and the procedure repeated.

Table 19 shows the results in terms of accuracy. Using the Self-Training method, we were able to improve the accuracy on predicting stress to 71.73% (+4.20%) and we were able to reduce the number of unlabelled data from 20% to 6%. In terms of gender, the results show that the *Male* achieved better accuracy 72% (*Precision*: 73.5%; *Recall*: 78.5%) for supervised approach and 76.4% (*Precision*: 73.5%; *Recall*: 78.5%) for SSL, in contrast to *Female* with 59.8% (*Precision*: 59.0%; *Recall*: 60.0%) for supervised and 64.8% (*Precision*: 62.0%; *Recall*: 65.0%) for SSL approach.

In the Table 20, we applied in our data set the *Variable Importance* (VI) and *Weight by Tree Importance* (WTI). VI evaluates the percentage of trees with a particular attribute

while WTI calculates the weight of the attributes by analyzing the split points of all the decision tree models and the attributes with higher weight are considered more relevant and important [35]. In this section, we show that simple models can be generated to predict stress levels with around 70% of accuracy. Unsurprisingly, most of the models used the relevant features identified in the previous section, whereas a slight improvement in the predictive performance can be achieved with a simple semi-supervised learning algorithm.

6 Conclusions

Stress at work is an issue that can affect employees' health, organisational performance, and reduce the quality of life in general. Assessing stress is a complex issue, in particular when relying on non obtrusive approaches.

In this work, we presented an extensive analysis based on real data from 30 users in two organisations related to stress using information derived from smartphones. We contrasted objective variables, obtained from smartphones, such as physical activity, location, social interaction and social activity, with respect to perceived stress levels, considering several demographics (gender, age, education and marital status). We presented classification results using decision trees, including the incorporation of semi-supervised learning methods.

A summary of the most important findings in this study are:

- There is correlation between smartphone objective data such as: location information (WiFi and Google Location data), social interaction, information from phone calls and SMS, with subjective data that represents the mood of the user (i.e., perceived level of stress).
- Our findings indeed confirm that it is feasible to detect perceived stress at work using smart phone sensed data
- Based on smartphone data it is possible to predict perceived stress with relatively high accuracy (above 70%)

Based on the results from the field studies, we can observe the following:

- There is clearly a moderate to high perceived stress in most of the employees.
- Overall, lower perceived stress was associated with higher physical activity, while high levels of perceived stress were associated with lower physical activity.
- Our results suggests that the more social the subject is the more stressed they become. While there is a negative correlation between social activity as measured by duration of calls and stress.
- It appears that women tend to present higher percentage levels of perceived stress. Whether this has to do with a biased finding due to our small sample size or to a more profound reason related to gender, would require further and deeper studies.
- Perceived stress varies among companies and this could be related to their working conditions. Identifying working conditions on companies with low levels of stress could help to establish better working policies to reduce stress among employees.

We believe our findings open new research perspectives to improve future monitoring systems towards a better understanding of stress at modern work. In particular, the results of our work can potentially impact occupational health as well

as management of human resources. High levels of stress negatively impact not only health but also productivity, increasing the risk of employee burnout. Therefore, assessing stress levels through methods such as ours, can become an additional instrument in occupational welfare decision making.

6.1 Future work

As future work we would like to continue with larger scale studies for the detection and prevention of stress at work. We also plan to analyse more in depth the decision trees obtained for each subject in order to obtain clusters of people who behave similarly; this could help us to build prediction models for new users with few data. Further, we want to test different levels of granularity for the time dimension to see whether stress patterns appear during different time intervals. Eventually, we aim to develop a smartphone app that can reliably detect stress levels in working environments.

References

- [1] E. Schneider, S. Copsey, and X. Irastorza, *OSH in figures: work-related musculoskeletal disorders in the EU-facts and figures*. Office for Official Publications of the European Communities, 2010.
- [2] NOISH, “Stress at work,” <http://www.cdc.gov/niosh/docs/99-101/>, 1999, accessed: 2015-AUG-29.
- [3] P. M. Bongers, C. R. de Winter, M. A. Kompier, and V. H. Hildebrandt, “Psychosocial factors at work and musculoskeletal disease,” *Scandinavian journal of work, environment & health*, pp. 297–312, 1993.
- [4] C. Maslach, W. B. Schaufeli, and M. P. Leiter, “Job burnout,” *Annual review of psychology*, vol. 52, no. 1, pp. 397–422, 2001.
- [5] P. Sanches, K. Höök, E. Vaara, C. Weymann, M. Bylund, P. Ferreira, N. Peira, and M. Sjölander, “Mind the body!: designing a mobile stress management application encouraging personal reflection,” in *DIS ’10 Proceedings of the 8th ACM Conference on Designing Interactive Systems*, Aarhus, Denmark, 2010, pp. 47–56.
- [6] X. Zhu, “Semi-supervised learning literature survey,” 2005.
- [7] J. Bakker, M. Pechenizkiy, and N. Sidorova, “What’s Your Current Stress Level? Detection of Stress Patterns from GSR Sensor Data,” in *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, Vancouver, Canada, 2011, pp. 573–580.
- [8] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, “Physiological signals based human emotion recognition: a review,” in *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*. Penang, Malaysia: IEEE, 2011, pp. 410–415.
- [9] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, “Towards mental stress detection using wearable physiological sensors,” in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 1798–1801.
- [10] T. Roh, K. Bong, S. Hong, H. Cho, and H.-J. Yoo, “Wearable mental-health monitoring platform with independent component analysis and nonlinear chaotic analysis,” in *Engineering in Medicine and Biology Society (EMBC)*, 2012 Annual International Conference of the IEEE, San Diego, CA, USA, 2012, pp. 4541–4544.
- [11] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury, “StressSense: Detecting stress in unconstrained acoustic environments using smartphones,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp ’12. New York, NY, USA: ACM, 2012, pp. 351–360. [Online]. Available: <http://doi.acm.org/10.1145/2370216.2370270>
- [12] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, “Activity recognition and monitoring using multiple sensors on different body positions,” in *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. International Workshop on*, Cambridge, MA, USA, 2006.
- [13] E. Garcia-Ceja, V. Osmani, and O. Mayora, “Automatic stress detection in working environments from smartphones’ accelerometer data: A first step,” *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1053–1060, July 2016.
- [14] A. Maxhuni, P. Hernandez-Leal, L. E. Sucar, V. Osmani, E. F. Morales, and O. Mayora, “Stress modelling and prediction in presence of scarce data,” *Journal of biomedical informatics*, vol. 63, pp. 344–356, 2016.
- [15] R. Likamwa, Y. Liu, N. D. Lane, and L. Zhong, “Moodscope: building a mood sensor from smartphone usage patterns,” in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, Taipei, Taiwan, 2013, pp. 389–402.
- [16] G. Bauer and P. Lukowicz, “Can smartphones detect stress-related changes in the behaviour of individuals?” in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, Lugano, Switzerland, 2012, pp. 423–426.
- [17] A. Sano and R. W. Picard, “Stress recognition using wearable sensors and mobile phones,” in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, Geneva, Switzerland, 2013, pp. 671–676.
- [18] A. Sano, A. J. Phillips, Z. Y. Amy, A. W. McHill, S. Taylor, N. Jaques, C. A. Czeisler, E. B. Klerman, and R. W. Picard, “Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones,” in *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on*. IEEE, 2015, pp. 1–6.
- [19] A. Muaremi, B. Arnrich, and G. Tröster, “Towards measuring stress with smartphones and wearable devices during workday and sleep,” *BioNanoScience*, vol. 3, no. 2, pp. 172–183, 2013.
- [20] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, and A. S. Pentland, “Daily stress recognition from mobile phone data, weather conditions and individual traits,” in *Proceedings of the ACM International Conference on Multimedia*, Orlando, Florida, USA, 2014, pp. 477–486.
- [21] E. Demerouti and A. B. Bakker, “The oldenburg burnout inventory: A good alternative to measure burnout and engagement,” *Handbook of stress and burnout in health care*. Hauppauge, NY: Nova Science, 2008.
- [22] D. M. McNair, M. Lorr, and L. F. Droppleman, *Profile of*

- mood states.* Univ., 1971.
- [23] FUNF, "Open sensing framework-funf," <http://funf.org/about.html>, 2015, accessed: 2015-AUG-29.
 - [24] D. Birant and A. Kut, "St-dbscan: An algorithm for clustering spatial-temporal data," *Data & Knowledge Engineering*, vol. 60, no. 1, pp. 208–221, 2007.
 - [25] C. Robusto, "The cosine-haversine formula," *American Mathematical Monthly*, pp. 38–40, 1957.
 - [26] P. Hedelin and D. Huber, "Pitch period determination of aperiodic speech signals," in *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*. Albuquerque, NM, USA: IEEE, 1990, pp. 361–364.
 - [27] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.
 - [28] I. Carreras, A. Matic, P. Saar, and V. Osmani, "Comm2sense: Detecting proximity through smartphones," in *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*. IEEE, 2012, pp. 253–258.
 - [29] R. S. Lutz, M. A. Stults-Kolehmainen, and J. B. Bartholomew, "Exercise caution when stressed: stages of change and the stress-exercise participation relationship," *Psychology of Sport and Exercise*, vol. 11, no. 6, pp. 560–567, 2010.
 - [30] E. Demerouti, A. B. Bakker, I. Vardakou, and A. Kandas, "The convergent validity of two burnout instruments: A multitrait-multimethod analysis," *European Journal of Psychological Assessment*, vol. 19, no. 1, p. 12, 2003.
 - [31] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.
 - [32] R. Ferdous, V. Osmani, J. B. Márquez, and O. Mayora, "Investigating correlation between verbal interactions and perceived stress," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 1612–1615.
 - [33] J. R. Quinlan, *C4. 5: programs for machine learning*. Morgan Kaufmann, 1993.
 - [34] I. Triguero, S. García, and F. Herrera, "Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study," *Knowledge and Information Systems*, vol. 42, no. 2, pp. 245–284, 2015.
 - [35] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, no. 1, p. 213, 2009.



Pablo Hernandez-Leal Pablo obtained his Ph.D. in 2015 from 'Instituto Nacional de Astrofísica, Óptica y Electrónica' (INAOE) in Mexico. His main research interest is machine learning, in particular in reinforcement learning and multi-agent systems.



Eduardo Morales is a Senior Researcher at the 'Instituto Nacional de Astrofísica, Óptica y Electrónica' (INAOE) in Mexico, where he conducts research in machine learning and robotics.



L. Enrique Sucar is Senior Research Scientist at INAOE, Puebla, Mexico. He has a Ph.D. in Computing from Imperial College and is Member of the National Research System, the Mexican Science Academy, Senior Member of the IEEE; and received in 2016 the National Science Prize.



Venet Osmani is a Senior Researcher at Fondazione Bruno Kessler (FBK) and Lecturer at University of Trento, Italy. His research is primarily focused on computational modelling of patient trajectories, based on analysis of Electronic Health Records using machine learning methods.



Oscar Mayora is senior researcher at FBK High Impact Initiative on Health and Wellbeing. He is founder of Pervasive Health Conference series and has coordinated a number of European Projects in the topic of digital health.



Alban Maxhuni obtained his Ph.D. in 2017 from University of Trento (DISI) in Italy. His research interests include biomedical signal processing, machine learning and pattern recognition.